

SAR Simple And Retrieval mit dem Siemens Texteditor EDT

Mohler, Peter Ph.

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Mohler, P. P. (1982). *SAR Simple And Retrieval mit dem Siemens Texteditor EDT*. (ZUMA-Arbeitsbericht, 1982/13). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205394>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

S A R

Simple And Retrieval

mit dem Siemens Texteditor EDT

Peter Ph. Mohler

ZUMA Arbeitspapier

Nr. 13/82

Juli 1982

ZUMA

B 2,1

68 Mannheim

1. Einleitung

Textverarbeitung mit Computern, gleich ob es sich um computerunterstützte Inhaltsanalyse oder eher qualitativ ausgerichtete Forschungsansätze handelt, ist oftmals mit "Information Retrieval" (Informations-Rückgewinnung) verbunden. Das Suchen bestimmter Begriffe oder Namen (Zeichenkette) in einer Textdatei ist die einfachste Art des Information Retrieval.

Faktisch alle Textmanipulationsprogramme (Editoren) enthalten Kommandos (Funktionen, Optionen), mit denen der Benutzer Zeichenketten in der Datei suchen und sich auf dem Bildschirm oder Drucker ausgeben lassen kann. 'Zeichenkette' ist jede Folge von Zeichen, also auch mehrere Worte hintereinander oder auch nur ein einzelnes Zeichen.

Mit diesen Funktionen kann der Benutzer interaktiv einzelne Zeichenketten rückgewinnen. Oft ist aber die Suche nach nur einer Zeichenkette unbefriedigend, weil eben die Kooccurrenz, das gemeinsame Auftreten von Begriffen, einer der wichtigsten Analysepunkte in der Textverarbeitung ist. Textanalyseprogramme wie TEXTPACK V oder OCP enthalten Funktionen, die auch bedingten Information Retrieval erlauben. Allerdings sind diese Programme nicht interaktiv verwendbar und die Daten müssen vor dem Suchlauf für diese Programme aufbereitet sein. Schließlich muß der Benutzer noch die Kommandosprache des jeweiligen Systems lernen. Für kleinere Suchläufe oder kurzfristige Analysen bedeutet dies einen wesentlichen Zeitverzug.

2. SAR mit dem Siemens-EDT

Benutzer, die über ein programmiertes Text-manipulationsprogramm, wie z.B. Siemens-EDT, verfügen, können dieses auch für einfaches bedingtes Suchen verwenden. Die Bedingungen können der Form nach logische ODER- und UND-Bedingungen sein.

Der Einfachheit halber beschränke ich mich auf den Fall der UND-Verbindung. UND-Verbindung bedeutet, daß ein Textteil nur dann vom Programm ausgesucht wird, wenn in ihm zwei oder mehrere Zeichenketten zugleich vorhanden sind. Also z.B. alle Zeilen einer Zeitung, in der Schmidt und Genscher vorhanden sind.*

Die Anweisung an das Programm lautet also:

1. nimm die erste Zeile des Textes und
suche darin die Zeichenkette 'Schmidt'
2. falls dies nicht zutrifft: gehe zur nächsten Zeile; sonst:
3. suche in derselben Zeile 'Genscher'
4. falls dies nicht zutrifft: gehe zur nächsten Zeile, sonst:
5. gebe diese Zeile aus (Bildschirm, Drucker)
6. fahre fort, d.h. gehe wieder zu 1 und nimm dann die nächste Zeile.

*Die Beschränkung auf Zeilen ist nicht notwendig, erfordert aber ein komplexeres Programm

Für den Siemens-EDT lautet eine solche Anweisung:

Sp.1
1. @ ON * FIND 'Schmidt'
2. @ IF.FALSE. GO TO 6
3. @ ON * FIND 'Genscher'
4. @ IF.FALSE. GO TO 6
5. @ PRINT *
6. @ CONTINUE

Diese Anweisung wird in eine sogenannte 'Prozedur' des EDT geschrieben und ist nichts anderes als ein kleines Programm, mit dem die Textdatei manipuliert wird.

Um eine solche Prozedur anzulegen, gibt man im EDT das Kommando

@ PROC 1	(es gibt bis zu 21 Prozeduren für eine EDT-Sitzung, die mit PROC 1, PROC 2 ... angelegt werden können.
* @ ON * FIND.	
...	
@ CONTINUE	
@ END	

'1' ist dann der 'Name' der Prozedur)
das Kommando ' END' schließt die Eingabe des Programms ab.**

NB. *die doppelte Angabe von @@ ist notwendig, damit der EDT diese Zeile als Eingabe eines Textes und nicht als Kommando interpretiert.

**eine Prozedur kann wie jede Textdatei geändert, erweitert etc. werden; vgl. dazu die einschlägigen Stellen im EDT-Handbuch.

Jetzt kann man, wenn man es noch nicht getan hat, die Textdatei in den Editor holen:

@GET 'TEXTDATEI'

und mit

@DO 1 * = 1, \$, 1

das Suchprogramm starten.

Es gelten folgende Konventionen für das

@DO-Kommando:

* = 1 ist die Angabe, daß die erste zu durchsuchende Zeile gleich der ersten Textzeile ist; \$ besagt, daß bis zur letzten Textzeile gesucht werden soll und die letzte '1' besagt, daß in ganzzahligen Sprüngen von Zeile 1 zu 2 zu 3 vorgegangen werden soll.*

* Der EDT kennt auch Zeilen mit Dezimalnummerierung, also Zeile 1.01, dann würde die Schrittweite nicht 1 sondern .01 betragen.

3. Erweiterungen der Prozedur

a) Drei und mehr Suchbegriffe

Sollen drei Zeichenketten berücksichtigt werden, so werden in der Prozedur zwei weitere Zeilen nach d) eingefügt, nämlich

4,1 @ON * FIND 'DRITTEN BEGRIFF'

4,2 @IF.FALSE. GO TO 8

und alle GO TO-Anweisungen lauten jetzt 8
(8. Zeile des Programms → CONTINUE;
d.h. in 2., 4. und 4.2 muß '8' stehen.

b) Erleichterung des Auswechselns von Suchbegriffen

In einem weiteren Suchlauf, bei dem z.B.
'Genscher' gegen 'Brandt' ausgetauscht werden
soll, muß der Benutzer mit

@PROC 1

wieder in die Prozedur 1 gehen und die Zeile 3
entsprechend ändern.

Dies kann er sich leichter machen, wenn er
statt der expliziten Nennung der Begriffe
Variablen einsetzt. EDT kennt 20 sogenannte
'Zeichenfolgevariablen' (#S1 - #S20), in denen
Zeichenketten abgelegt (gespeichert) werden
können.

Die Zeilen 1. und 3. heißen dann

1. @ON * FIND #S1

2. @ON * FIND #S2

bevor die Prozedur mit

@DO 1 * = 1, \$, 1

gestartet wird, muß der Benutzer angeben,
für welche Zeichenketten #S1 und #S2 stehen
sollen. Dies geschieht mit der Anweisung

```
@SET #S1 = 'Schmidt'  
@SET #S2 = 'Genscher'
```

beim nächsten Suchlauf, in dem 'Genscher' durch 'Brandt' ersetzt werden soll, heißt dann die Kommandoabfolge:

```
@SET S2 = 'Brandt'  
@DO 1 * = 1, $ , 1
```

man braucht jetzt also nicht mehr in die Prozedur zurückzugehen.

c) Spaltenweises Suchen

Für Benutzer, die in bestimmten Spalten einer jeden Zeile bestimmte Informationen verschlüsselt haben (Fixed Format), kann die Suche auch auf Spalten innerhalb einer Zeile begrenzt werden.*

Dazu muß die Angabe für die Zeile * um die Angabe der jeweiligen Spalten ergänzt werden: *: a-b
a und b sind dann Ziffern und bezeichnen die Spalte ab der und bis zu der gesucht werden soll.

Hier ein Beispiel:

In der Studie 'Psychoanalytische Ambulanz' wurde für jede Zeile der Textdatei angegeben, welches Thema im Text abgehandelt wird. Die Angabe 'Thema' wurde als vierspaltiger Code in die Spalten 7-10 einer jeden Zeile geschrieben, der eigentliche Text steht in Spalte 12-80. Der Code für 'Diagnose' sei DIAG und es werden alle Diagnosen gesucht, in denen die Zeichenkette 'Psychoanalyse' vorkommen.

* Diese Version ist auch besonders für numerische Dateien geeignet!

Dann lautet die Suchprozedur:

```
@@ON *: 7-10 FIND 'DIAG'  
@@IF.FALSE. GO TO 6  
@@ON *: 12-80 FIND 'PSYCHOANALYSE'  
@@IF.FALSE. GO TO 6  
@@P *  
@@CONTINUE
```

Durch die Beschränkung auf bestimmte Spaltenbereiche wird natürlich auch die Suchzeit verringert.

- d) Parameterübergabe (Schlüsselwörter) außerhalb der unter b. ist beschrieben, wie man mit dem SET-Kommando die Schlüsselwörter beim Ablauf der Prozedur eingeben kann. Der Vollständigkeit halber sei hier noch auf die Möglichkeit des EDT verwiesen, sogenannte Parameter, d.h. Leerstellen, in der Prozedur vorzusehen, die dann innerhalb eines erweiterten DO-Kommandos spezifiziert werden. Genaueres ist im EDT-Handbuch nachzulesen.

e) Sichern der Prozedur

Man kann die Prozeduren wie jede normale Datei mit einem SAVE-Kommando abspeichern und später mit einem GET-Kommando wieder laden:
z.B.:

```
@PROC 1  
.....  
....  
@SAVE 'PM.PROZEDUR.13'  
@END  
@PROC 14  
@GET 'PM.PROZEDUR.13'  
.....
```

f) Protokollieren der Suchläufe

Die interaktiven Suchläufe können auf sehr einfache Weise protokolliert werden:

In BS 2000, dem Betriebssystem des Mannheimer Rechners, definiert man am besten zuerst eine Ausgabedatei mit dem Kommando

```
/SYSFILE SYSLIST = name der ausgabedatei
```

dann weist man das Betriebssystem an, alle Ein- und Ausgaben am Bildschirm in die 'ausgabedatei' zu protokollieren:

```
/OPTION MSG=FH
```

ruft man jetzt den Editor auf, so werden u.a. alle erfolgreichen Suchvorgänge in dieser Datei protokolliert.

Ist man mit der Suche fertig und hat das Programm EDT abgeschlossen (Kommando \odot H), so wird die Protokolldatei mit folgenden Kommandos ausgedruckt:

```
/OPTION MSG=F          - keine Protokollierung mehr  
                        der Ein-Ausgabe
```

```
/SYSFILE SYSLST = (PRIMARY)  
                  Rückschalten auf System-  
                  Druckfile
```

```
/PR   ausgabedatei, T = STATIONZ, LINES = 60*
```

* T = STATIONZ - Ausgabe bei ZUMA
 LINES = 60 - maximal 60 Zeilen pro Seite,
 wegen der "Schönheit" des
 Druckbildes.

NB. Alle /- Kommandos sind von Änderungen des Betriebssystems abhängig. Stand Juli 1982!

